Ecological Informatics 90 (2025) 103231

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Flexible hybrid edge computing IoT architecture for low-cost bird songs detection system

Francisco A. Delgado-Rajó *, Carlos M. Travieso-Gonzalez

Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC) Science and Technology Park, Edificio Polivalente II, 2ª planta, C/ Practicante Ignacio Rodríguez, s/n, 35017 Las Palmas de Gran Canaria, Spain

A R T I C L E I N F O

Keywords:
Birds song recognition
Edge computing
Low-power wide-area networks
Biodiversity
Internet of things

A B S T R A C T

The monitoring of bird populations provides valuable insights into biodiversity variations and their correlation with environmental changes. This study proposes a flexible hybrid edge computing IoT architecture for a low-cost bird song detection system. The system integrates low-power microcomputers, such as Raspberry Pi, equipped with USB microphones, LoRa modules, and Wi-Fi for seamless operation across rural and urban environments. By utilizing deep learning techniques, including convolutional neural networks (CNNs) trained on bird song datasets, the system performs real-time species detection at the edge, minimizing the need for high-bandwidth transmission. Nodes dynamically select communication technologies based on availability, sending data to an IoT analytics platform. Field deployments demonstrate the system's efficiency, interoperability, and adaptability for biodiversity monitoring, particularly in remote areas with limited connectivity. This architecture addresses the challenges of real-time species detection while ensuring low cost, scalability, and energy efficiency. The main advantage is that devices can operate in areas without mobile coverage, as they only transmit the detection signal. This results in significant bandwidth savings, since the processing is carried out at the edge.

1. Introduction

In recent years, the observation of biodiversity has become more prominent than ever, especially its changes over time and space due to different factors. Their study makes it possible to analyse and predict the effects of these factors on different ecosystems. Climate change becomes relevant, affecting the desertification of large areas, the variation of the migrations of many species or the disappearance of others in different areas. It is also important to be able to detect the influence of certain invasive species on local bird's populations.

It is, therefore, crucial to monitor the correlation between these factors and variations in biodiversity indices. For this analysis, different variables, known as Essential biodiversity variables (EBV's) (Jetz et al., 2019), can be used for mapping and monitoring populations of multiple species. In this study, it is advocated to create a three-dimensional map (species, space, time) to reflect the presence or absence of a given species in each place or at a certain time (Andrewartha and Birch, 1954; Bell, 2003; Jetz et al., 2012). Data can come from Incidental observations, inventories (Guralnick et al., 2017) or expert data collection.

According to the "State of the World's birds" in the Annual Review of Environment and Resources (Lees et al., 2022), birds can be considered an important indicator of how biodiversity indices vary (Schmeller et al., 2012). According to (Lees et al., 2022), 90 % of extinct bird species are endemic to islands. Specifically, in the case of the islands of Macaronesia, 32 species have disappeared since 1900.

This is why, in recent years, significant effort has been made to monitor the populations of different bird species in any area and for long periods of time. To do this, various strategies are used that can cover large areas of land. In the case of the observation of migratory movements, for example, there are multiple studies that use radar techniques (Liechti et al., 2019; Nilsson et al., 2018) to monitor them. Another strategy is based on the observation of nests or artificial feeders visited by birds, based on RFID (Pereira et al., 2023; Youngblood, 2019), or on automatic systems with cameras for video recordings or photographs (Binta Islam et al., 2023; Akçay et al., 2020). Evidently, the improvement in sensor monitoring techniques has grown exponentially due to their lower cost, lower consumption, or the incorporation of new specific communication techniques (La Sorte et al., 2018). The latter also makes it possible to relate the data on the existence or non-existence of a species to the climatic variables of the environment, for example. Another form of monitoring is based on Citizen Science (Ferdoush and Li, 2014; Rovithis et al., 2021; Sullivan et al., 2009), based on the

* Corresponding author.
E-mail address: paco.rajo@ulpgc.es (F.A. Delgado-Rajó).

https://doi.org/10.1016/j.ecoinf.2025.103231
Received 10 January 2025; Received in revised form 14 April 2025; Accepted 24 May 2025
Available online 27 May 2025
1574-9541/

collection of data by volunteers who use mobile applications to acquire the data that is collected by an organization. This becomes more relevant in urban or suburban environments where there is 4G or 5G or Wi-Fi connectivity. This allows the automatic detection of species through photos or audio recordings, as well as audio recordings that can be used for training automatic detection systems afterward.

In recent years, the emergence of deep learning techniques has enabled systems to automatically recognize species through the sounds they produce. An experienced observer can detect the presence of an individual of a given species by the sound of its song. In the same way, many works have been developed that allow this detection to occur automatically. For this purpose, sound recording equipment is installed in the usual habitats and sends audio files corresponding to hours of recording in which the songs of various species appear (LeBien et al., 2020; Ventura et al., 2015) or could transmit in real time the sound collected by the audio equipment (Aide et al., 2013; European Commission, 2011). This transmission of audio in real time or audio files for further processing requires relatively high bit rates and higher consumption than the transmission of sensor data using specific IoT communications technologies. This is especially critical in rural areas where there is no 4G/5G or Wi-Fi coverage. The incorporation of new low-cost devices such as microprocessors or Low-Power microcomputers into the market makes it possible to combine sensor data collection with processing and data transmission capacity. The deployment of nodes along the terrain that allows audio recording (LeBien et al., 2020) can be complemented with additional nodes that, through processing, allow the identification of certain species of birds in real time and could send an alert in the event of the appearance of any of them. In this way, large areas are monitored continuously over time. The latter is the main objective of this work.

1.1. Deep learning techniques in edge computing

Various methods have been employed to detect species by identifying their songs. For example, (Daidai Liu et al., 2025) provide an overview of the most used techniques for species recognition, covering approaches for segmenting song fragments as well as deep learning-based recognition methods. These processes often follow a four-step pipeline: dataset acquisition, preprocessing, feature extraction, and classification. Traditionally, most studies perform the last three stages on powerful servers.

However, the potential of edge computing remains largely unexplored in this field. Edge computing enables real-time data processing on low-power devices such as microcomputers, eliminating the need for constant connectivity to centralized servers. This approach offers significant advantages, including reduced latency, enhanced data privacy, and lower energy consumption, making it particularly suitable for ecological monitoring in remote environments.

Acoustic identification is widely recognized as more effective than visual methods for species recognition (Stowell et al., 2016). While early works in this field relied on deterministic techniques and preprocessing-heavy pipelines (Jančovič and Kküer, 2011; Priyadarshani et al., 2018; Wielgat et al., 2011), more recent studies have embraced artificial intelligence (AI) techniques, such as convolutional neural networks (CNNs), for species-specific song recognition. Yet, these advancements have been primarily implemented on high-performance computing platforms, leaving a gap in the development of lightweight solutions tailored to edge devices.

Recent research in acoustic monitoring has demonstrated the potential of deep learning for biodiversity studies. For instance, (Stefan Kahl et al., 2021; Noumida and Rajan, 2021) developed BirdNET, a solution leveraging CNNs for species identification from audio recordings, showcasing the utility of spectrogram analysis. Similarly, (Goitia-Urdiain et al., 2024) highlighted biases in software-dependent recognition of bird songs, emphasizing the importance of robust methodologies to avoid misinterpretations of bird abundance and vocal activity.

Spectrogram processing using CNNs is particularly well-suited for edge-based systems, given its compatibility with compressed audio inputs and its potential for high accuracy. For example, (B. Chandu et al., 2020) highlights the potential of lightweight architectures for field deployment. Additionally, (Velasco-Montero et al., 2024) underscores the importance of continual learning frameworks, which could be adapted to edge devices for long-term ecological studies.

Recently, studies have emerged (Arowolo et al., 2024) that integrate Artificial Intelligence with Internet of Things (IoT) devices and communication techniques. This approach brings processing closer to sensor devices, thus reducing network load and system latency. By leveraging edge computing, this study aims to optimize spectrogram-based bird song recognition for low-power devices, striking a balance between computational efficiency and classification accuracy. This not only enhances the portability of acoustic monitoring systems but also enables real-time applications in remote and resource-constrained environments.

Recordings, usually obtained in wild environments, are used to train the neural network (Andreassen et al., 2014; LeBien et al., 2020) which involves an automatic call detection and segmentation process to isolate the recording fragments that contain actual bird songs. This process is especially necessary when the percentage of recording time with singing is very small. In any case, this process is more complicated in noisy environments, so it is often followed by an additional manual process performed by experts for segmentation prior to the training process of the systems (Rocha et al., 2015; Selin et al., 2007). In most of the cited studies, either the audio segments obtained through recording devices must be transmitted, or these segments require post-processing in a local environment. The main drawback is that, except for a few cases, real-time detection and classification using Edge computing is not implemented (Guma et al., 2025).

Other studies emphasize the use of low-power devices to implement an initial detection phase, followed by a more precise species recognition phase (Cinkler et al., 2021). In such systems, the first phase extracts features and filters the input, while the second phase employs a pre-trained neural network trained on bird sound datasets for classification. For instance, (Ester Vidaña-Vila et al., 2020) propose a two-layer classification approach. The first layer determines whether a bird sound is present. If the first layer identifies an acoustic sample as a bird sound, the second layer is activated to classify the specific bird species and sound type. In these two studies, it is ultimately necessary to retransmit the audio segment, losing the ability to perform real-time detection and classification with low latency.

In the present study, this process is carried out on-site. This approach eliminates the need to transmit audio to a central node. Instead, only the alerts generated by the detection and classification process are transmitted, significantly reducing the required bitrate. This enables the use of alternative long-range, low-power communication techniques, as detailed later.

1.2. Microcomputers

There are currently multiple low-cost and low-power devices that have moderate processing capacity. Among them, the Raspberry series (McBride and Courter, 2019) allows the implementation of artificial intelligence systems, in addition to managing wireless communications through compatible communications modules of different technologies and communicating with different sensors using their input/output ports (Ferdoush and Li, 2014). Additionally, as in (Travieso et al., 2021), USB cameras or microphones can be incorporated to collect images or sounds. One application of this device in the detection of bird songs could be the scheduling of automatic recording in wild environments depending on the presence or absence of a specific bird, for example. But, the highlight of this device compared to Arduino (Nayyar and Puri, 2016), for example, is the processing capacity and the large number of software libraries that facilitate its programming along with the easy interconnection of peripherals (cameras, microphones o communications modules). In addition, it allows programming in C++, Python or Matlab to enable the implementation of Internet of Things (IoT) systems with artificial intelligence autonomously, respecting the corresponding limitations of processing capacity inherent to these devices (Ariyanto et al., 2019; Weychan et al., 2015).

1.3. Communications networks

The systems have their own communications module based on 802.11 (Wi-Fi) embedded in the device. In certain urban scenarios, for example, an access point of this type can be found nearby, so the connection will be direct with the requirements of this type of network in terms of high bit rate, low latency and with the consequent restriction of range of less than 100 m, although in the latest versions (802.11 ac, 802.11n and 802.11ax) has been increased. In rural or suburban settings, however, a nearby hotspot is often not available. In these cases, the use of LPWAN technologies is required, in any of its options (Chaudhari et al., 2020; Delgado-Rajo and Alvarado Ramírez, 2024).

Among these technologies, there are those based on a proprietary network architecture (LoRa, LoRaWAN or SigFox), which require an interconnection gateway with the IP network and allow ranges of tens of kilometers thanks to their spread spectrum modulation (Vangelista, 2017). The main constraint of these systems is the basic transmission rate of less than 20 kbps and the limited size of the data frame. They are mainly used for sensor networks or small alerts over long distances where latency or even error rates are not critical. In birdsong detection systems such as those discussed here, they would not be able, for example, to transmit audio segments to the Gateway. Other options that allow a higher bitrate are those based on cellular networks (NB-IoT, LTE-M or Cat-M1) they also have a long-range and relatively low consumption, but they need 3G, 4G or 5G coverage depending on which one is chosen. There are other solutions based on interoperability between various IoT communications technologies (Delgado-Rajo et al., 2020; García-Martín and Torralba, 2021) that allow combining the characteristics of these according to the required bit rate and coverage, allowing the creation of clusters or Nano cells.

In this work, a flexible architecture that adapts to the scenario is proposed, enabling sound detection and recognition at the EDGE while alerting about the presence of each species, regardless of the communication technology used. In this way, the device connects to the cloud using one technology or another, depending on its location whether an urban or rural environment, and also facilitates the transmission of data from weather sensors at required points. This eliminates the need for a high bit rate, while allowing real-time processing near the sensors themselves. As a result, LPWAN techniques can be employed without the need for mobile coverage, unlike other works previously cited.

The main innovations proposed are:

a) The detection of bird songs is processed at the EDGE layer using low-cost and low-power devices. This eliminates the need to transmit audio segments, as done in other systems, and detects presence in the form of a percentage instead.
b) Despite the simplicity of the devices used, the detection results achieved are comparable to those found in the state of the art.
c) The final system offers flexibility in selecting the most suitable communication technology based on availability.
d) These nodes can be deployed in environments without mobile network coverage by utilizing LPWAN techniques, enabling real-time detection and classification.

2. Materials and methods

In this work, a deployment of different nodes managed by Raspberry Pi 3B+ has been carried out in different locations to check the interoperability of the different communication technologies of each node and the operation of the birdsong detection system in different environments and in real time. It is necessary to consider the limited processing and bitrate capacity of these nodes. In addition, the processing is done at the EDGE so high bit rates are not required and does not produce any overload on the network. Each node will be equipped with a USB microphone, a battery and a LoRa communications module, in addition to the Wi-Fi module embedded in the Raspberry.

2.1. Network architecture

The system architecture is shown in Fig. 1. As shown, it is a flexible network architecture that allows interoperability between two communications technologies. This is the main advantage of the system over others that require Wi-Fi coverage exclusively (Cinkler et al., 2021). Each Bird Songs Detection node (BSD) allows the choice between the two options depending on the availability or not of the Wi-Fi access point. If there is such an access point in the range, it connects using its SSID and password. If it does not exist, it sends the data from the detection of the chosen species through a LoRa connection to the nearest Gateway, which, in turn, retransmits it to the cloud using a wired or wireless IP connection to an IP router. This data is sent to a proprietary Backend or, in this case, an IoT analytics platform service like Thingspeak. In this way, the versatility of using the same type of node for different scenarios is facilitated with a simple configuration by the user in the deployment. A bridge is implemented in the gateway between the data coming from the LoRa network and the 802.11 network that allows the immediate publication of alarms in the Backend. The Gateway uses the Heltec LoRa wireless bridge with "Wi-Fi/ Bluetooth-LoRa" signals, ESP32, SX1276, which is compatible with the Arduino development environment with which the message forwarding code has been developed. The LoRa SX1276 chip that it implements uses a frequency of 868 MHz (Europe) that allows sending frames of 256 bytes with CRC and a maximum bit rate of 300Kbps. The −148 dBm sensitivity allows for long ranges in environments without too many obstacles.

The amount of information transmitted during each detection event is relatively small, resulting in low bandwidth requirements. Fig. 2 illustrates the format of the LoRa data frame transmitted by each node and provides an abstraction of the system architecture, highlighting the different layers. The detection process is carried out at the Edge layer, reducing the data load on the upper layers. The data frame fields include the node identifier, its location, the activation signals of the neural network output (integers ranging from 0 to 9) corresponding to each song, and a timestamp. It is important to note that this data is transmitted only when a species is detected based on a predefined activation signal threshold. In the case of Wi-Fi communication, the circuit embedded in the Raspberry Pi, along with its libraries, is directly used to publish the same data to Thingspeak, maintaining the same structure as for LoRa. Additionally, Fig. 2 also shows an abstraction of the employed Edge architecture, detailing the devices and physical components that

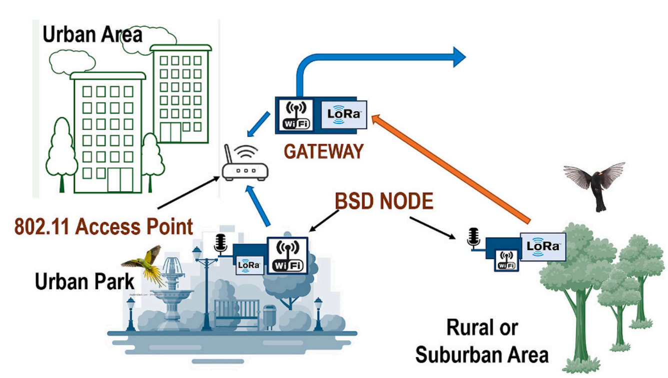

Fig. 1. System architecture.

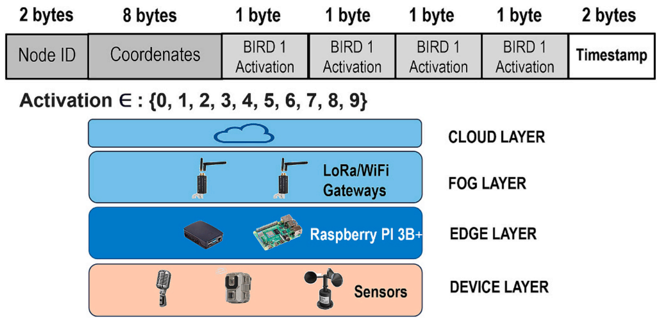

Fig. 2. LoRa's frame format and system layers.

make up each layer in this case.

2.2. Bird songs detection node implementation

As mentioned above, each node of this architecture is implemented using a Raspberry PI 3 B+ as its central core. In the block diagram of Fig. 4, it can be seen that the Raspberry itself contains the audio acquisition and subsequent processing blocks, which are used to obtain the images corresponding to the spectrograms of the 1 s audio segments that feed the trained neural network, as reflected in the previous section. In addition, a function that takes the activation signals of the neural network as inputs to detect the presence or not of the different species is implemented. It is important to note that, although the neural network is trained for the species mentioned above, depending on the location in each case (urban or rural) the species of interest in each case have been selected. These functions allow communication to be activated and data frames to be formed with the values of position, activation and moment of time when the data is produced, as well as writing data to the Raspberry's USB serial port for LoRa communication in case a Wi-Fi access point is not available. Both the sending of data using Wi-Fi and LoRa are carried out when a certain threshold is reached in the probability of detection of any of the species involved in order to avoid overloading the network or the data load at the Backend. In this case, a threshold value of 70 % probability has been selected.

The entire sound processing and detection system that is implemented in the Raspberry has been programmed using Matlab and Simulink (The MathWorks Inc, 2024) which allows the deployment of complete systems in this type of devices making use of the appropriate Deep Learning libraries, for example.

The Heltec LoRa wireless bridge module with "Wi-Fi/ bluetoothLloRa" (The same one used to implement the Gateway) is connected via an USB serial port just like the USB microphone. The whole set is powered by a Diymon 18,650 battery charging module that offers a power supply of 3 V/5 V and 1 Amp. Both writing to the serial port for LoRa communication or direct publishing via Wi-Fi occur only when there is a detection of one of the species. That is, when the level of probability of appearance of one of the species is higher than the fixed threshold.

2.3. Training process

The learning method selected in this work is a Convolutional Neural Network (CNN). This type of network is particularly effective for image processing tasks. In this instance, it is used to analyse spectrograms of bird songs.

The CNN consists of five simple convolutional layers. Each layer applies a series of filters to the input data, followed by a non-linear activation function. The purpose of these layers is to progressively extract higher-level features from the input data. Following the convolutional layers is a fully connected layer. This layer takes the high-level features extracted by the convolutional layers and uses them to make a final prediction. The outputs of the fully connected layer are the activation signals of each bird species, represented by the probabilities of each species' appearance.

The input to the network is a spectrogram image of a bird song. This image is obtained by applying a Mel-frequency spectrogram to audio fragments of bird songs (Carvalho and Gomes, 2023). These audio fragments are standardized to a duration of 1 s with a sampling rate of 16,000 samples per second. The spectrogram is calculated using 0.25 s frames duration for spectrum calculation and time hops of 0.01 s. The resulting spectrogram image has a resolution of 98 × 50 pixels, due to the use of a bank of 50 filters. The output of this layer is a set of feature maps that highlight these low-level features in the input images.

The second layer takes the feature maps from the first layer as its input. It applies another set of filters, which are typically more complex than those in the first layer. These filters are designed to detect higher-level features, such as textures and patterns, that are composed of the low-level features detected by the first layer. The output of the second layer is another set of feature maps that highlight these higher-level features.

The fully connected layer that follows the convolutional layers takes the high-level, abstract features extracted by the convolutional layers and uses them to make a final prediction. In the context of species classification, the outputs of the fully connected layer are the probabilities of the input belonging to each species.

The training data for the network comes from a database of bird songs, specifically from xeno-canto.org (xeno-canto, 2020). This database contains hours of recordings of up to 11,939 different species. The audio fragments used for training are carefully selected to contain the real song of the bird. A filtering process is used to segment the audio data and select fragments according to a signal threshold at the desired frequency as it can be seen in Fig. 3.

The species selected for this example are *the Heineken Eurasian Blackcap* (*Sylvia atricapilla heineken*), the *Canary Islands chiffchaff (Phylloscopus canariensis)*, the *Common blackbird (Turdus merula)*, the *Spanish Sparrow (Passer hispaniolensis)* and, finally, the *Rose-ringed parakeet (Psittacula krameri)*, the latter being a widespread invasive species in urban areas of cities in much of Europe. The aim is to see their influence on the other two endemic species of the Canary Islands. For each of the nodes, the network can be retrained using only the species that may be present depending on the area, since there is diversity between urban and rural areas. This makes the computational load of each node less heavy with the corresponding consumption savings. The trained network is the one used by the Simulink model implemented on the Raspberry for detection and recognition.

In summary, the five layers in the CNN architecture work together to transform the raw input data into a form that can be used for effective species classification. They do this by progressively extracting more complex and abstract features from the data, building up a hierarchical representation of the data that captures the most important aspects for the task at hand. This hierarchical feature extraction is one of the key strengths of CNNs, and it is what allows them to achieve state-of-the-art performance on a wide range of image classification tasks, including

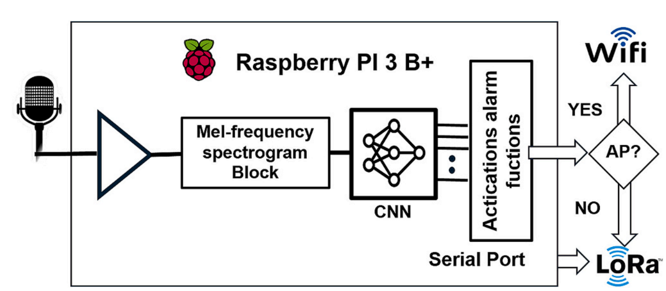

Fig. 3. BSD node blocks diagram. All blocks inside the rectangle are implemented in Simulink and deployed on the Raspberry PI.

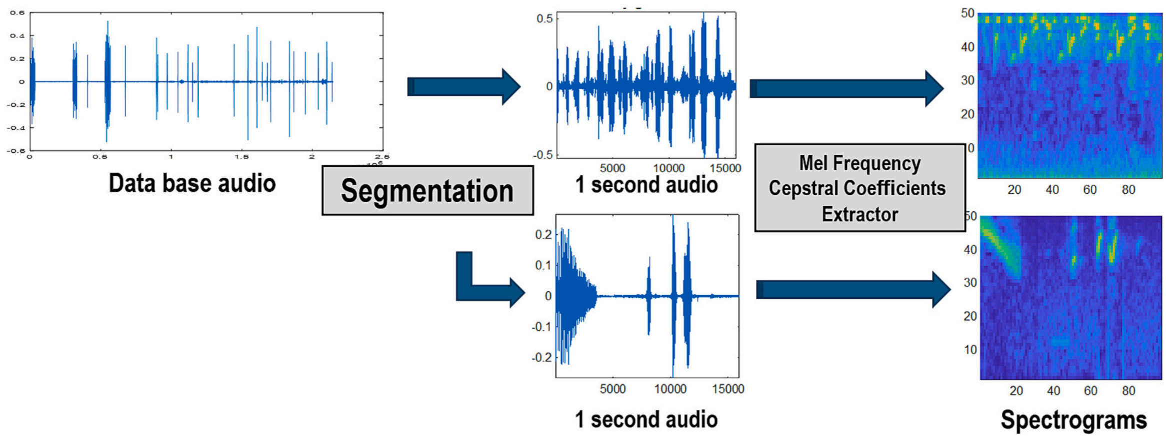

Fig. 4. Training process.

Fig. 5. Localization of the different test scenarios in Gran Canaria.

species classification.

2.4. Data analysis

For the verification of the system, as well as the communications systems, various scenarios with different characteristics have been used. On the one hand, measures have been taken in urban parks that had Wi-Fi connectivity, on the other hand, the system has been tested in a rural environment also with Wi-Fi connectivity and, finally, in a rural environment using a LoRa connection to a remote access point. For the tests carried out in these scenarios, the species that can be found in each one have been chosen. Fig. 5 shows the distribution of these scenarios. Scenarios 1 and 2 of the Fig. 5 correspond to urban parks and 3 to a rural area.

In each one, a complete node has been deployed with the different communication technologies integrated, making use of one or the other depending on availability. The nodes have been located anchored on poles or in trees during the testing period. These tests were carried out during various time slots in periods of approximately 1 h. Fig. 6 shows the implementation of the node with its protective housing. Since the goal is to test the interoperability of the system, the battery life is limited to 2 h maximum, so the devices were removed after the measurement periods. The results obtained are stored and visualized on the *Thingspeak* platform. The data obtained are the percentages of occurrence of each species over time, as well as the distribution of the total occurrences of the same during the total of all the measurement periods.

2.5. Performance evaluation metrics

To evaluate the effectiveness and robustness of the proposed methodology, a comprehensive performance analysis was conducted encompassing both class-specific and global metrics. The evaluation framework incorporates multiple performance indicators: for individual classes, the analysis includes confusion matrices, accuracy, recall, and F1-scores; at the system level, total accuracy and macro-average F1-score were computed to assess overall performance.

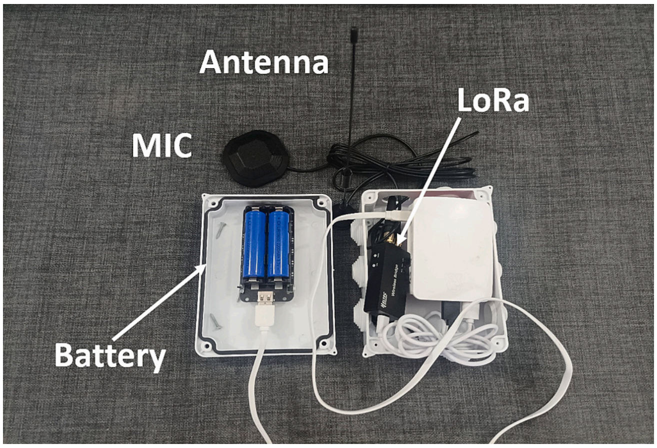

Fig. 6. Node implementation with Raspberry PI 3+. The LoRa module and the microphone are connected to the USB ports of the Raspberry Pi. The connection with the LoRa module is in serial mode

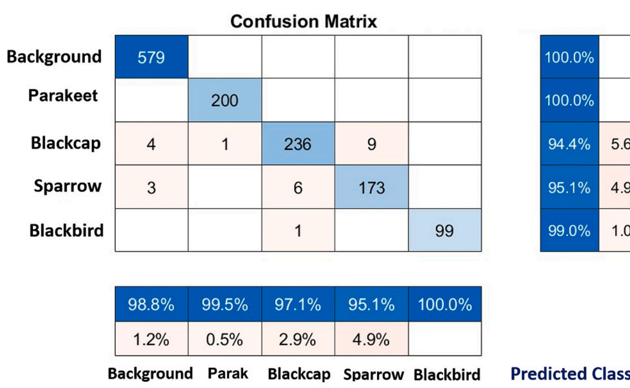

Fig. 7. Confusion matrix of species detection in Urban Parks.

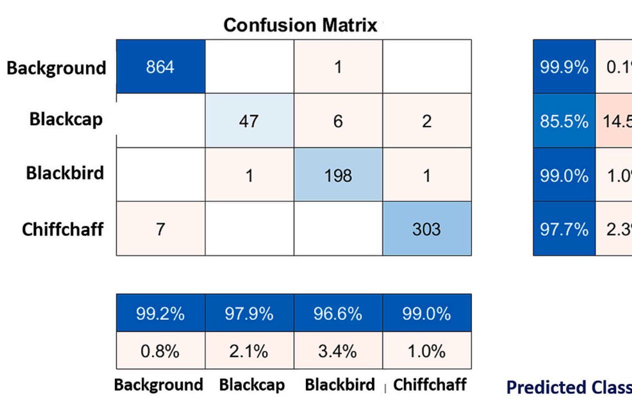

Fig. 8. Confusion matrix of species detection in Rural Area.

The confusion matrix methodology served as the foundational analytical tool for deriving these performance metrics. This evaluation framework is instrumental in classification tasks, providing a systematic quantification of model performance through the categorization of outcomes into four fundamental classifications: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

The classification outcomes are defined as follows:

- True Positives (TP) denote instances where avian vocalizations were accurately identified within the acoustic environment, with the model's species classification corresponding precisely to the ground truth annotations in the environmental recordings.
- False Positives (FP) encompass instances of misclassification where either ambient environmental acoustics were erroneously identified as avian vocalizations, or species were incorrectly classified, representing type I errors in the detection system.
- True Negatives (TN) represent correct classifications of non-avian acoustic events, demonstrating the model's capability to effectively discriminate between bird vocalizations and ambient environmental sounds in the acoustic landscape.
- False Negatives (FN) constitute instances where the model failed to detect verified avian vocalizations or incorrectly classified the species present in the recordings. These type II errors have significant implications for ecological monitoring and biodiversity assessment protocols.

Through the implementation of this comprehensive confusion matrix methodology, the study achieved a rigorous evaluation of the model's performance characteristics, facilitating a detailed analysis of its capabilities and limitations in species identification within complex acoustic environments. This analytical framework provides crucial insights into the model's efficacy in automated bioacoustics monitoring applications.

These metrics offer a quantitative assessment of the model's capacity to accurately classify instances as positive or negative, with a particular focus on distinguishing between background and bird types. The equations employed to compute each metric are detailed below:

$$Accurracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1 - score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{4}$$

The evaluation framework incorporates multiple complementary performance metrics to assess the model's classification capabilities. Accuracy quantifies the overall proportion of correct predictions across all classes, encompassing both avian vocalizations and environmental sounds. Precision measures the ratio of true positives (TP) among all positive predictions, evaluating the model's reliability in correctly identifying specific bird species when a detection is made. Specificity calculates the proportion of true negatives (TN) among all negative cases, assessing the system's capability to accurately discriminate background environmental sounds from bird vocalizations.

Recall (also termed Sensitivity, Hit Rate, or True Positive Rate [TPR]) quantifies the proportion of correctly identified bird vocalizations within each species class relative to the total number of actual occurrences in the acoustic dataset. This metric is mathematically expressed as the ratio of true positives to the sum of true positives and false negatives, providing crucial insights into the model's capability to detect specific avian species when they are present in environmental recordings. In the context of bioacoustics monitoring, recall serves as a particularly significant indicator, as it directly reflects the system's effectiveness in species detection and its robustness against false negative classifications that could impact biodiversity assessments.

The F1-score, computed as the harmonic mean between precision and recall, provides a balanced assessment of the model's performance by considering both false positives and false negatives, particularly valuable in ecological monitoring scenarios where both accurate species identification and minimal misclassification are crucial. These metrics collectively provide a comprehensive evaluation of the model's capacity to differentiate between various bird species and ambient acoustic

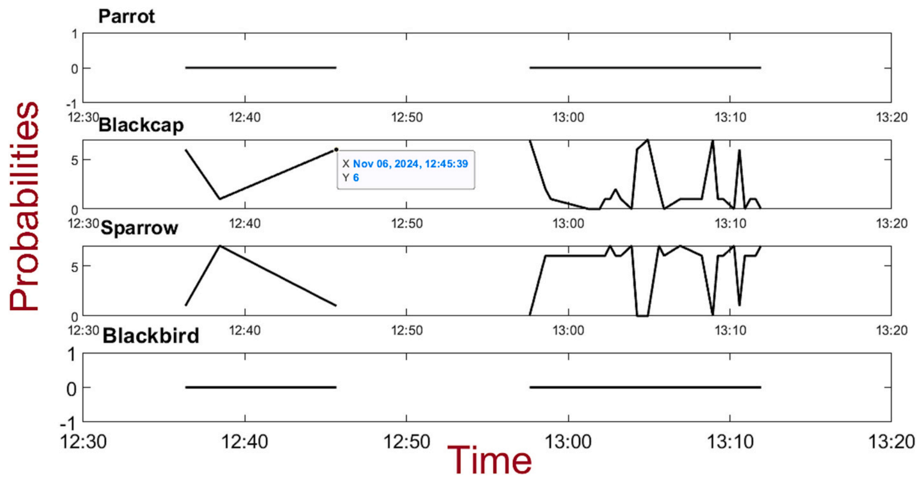

Fig. 9. Parrot, Blackcap, Sparrow, and Blackbird detections over a 50-min period at location1.

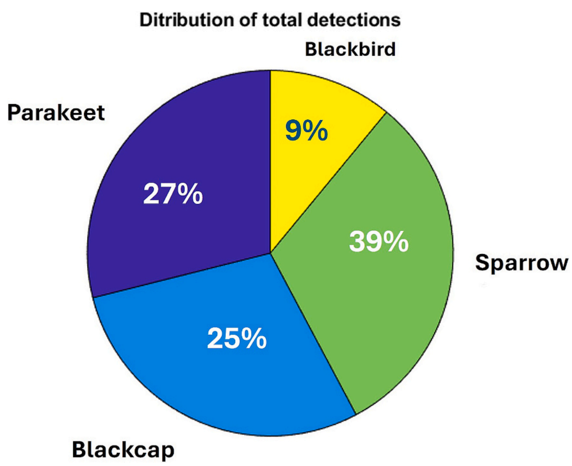

Fig. 10. Percentage of total detections over a period of 10 days at locations 1 and 2.

events in complex environmental recordings.

Total accuracy and macro-averaged F1-score were calculated as the arithmetic mean of the corresponding metrics for each class.

Accuracy refers to the percentage of correct predictions made by the model for both positive and negative events. Precision, on the other hand, measures the proportion of true positives (TP) among all cases predicted as positive, evaluating the model's accuracy in predicting events that occurred. Specificity calculates the proportion of true negatives (TN) among all negative cases, providing an assessment of the model's ability to correctly identify patients who do not require ICU admission. Lastly, the F1-score measures under the concept of harmonic mean, which is useful to find the best trade-off between the two quantities. They indicate the model's ability to correctly detect the type of bird and background.

3. Results

Firstly, as mentioned above, two neural networks have been trained: one for urban areas and the other for rural areas due to the differences between the species that we find between the two environments. In the first case (urban parks) the following species have been considered: Heineken Eurasian Blackcap (*Sylvia atricapilla heineken*), the Common blackbird *(Turdus merula),* the Spanish Sparrow *(Passer hispaniolensis)* and, finally, the Rose-ringed parakeet *(Psittacula krameri)*. The confusion matrix obtained is shown in Fig. 7. In the second case (Rural area) the following species have been considered: Heineken Eurasian Blackcap (*Sylvia atricapilla heineken*), the Canary Islands chiffchaff *(Phylloscopus canariensis*) and the *Common* blackbird *(Turdus merula)* whose confusion matrix is reflected in Fig. 8. On the left, the actual values are shown, and at the bottom, the percentage of correct and incorrect detections.

Table 1 shows the metric by classes and global metrics of the first ranking system. It shows a very high performance, with overall accuracy of 99.0 % and a macro-average F1-score of 0.977, indicating a very

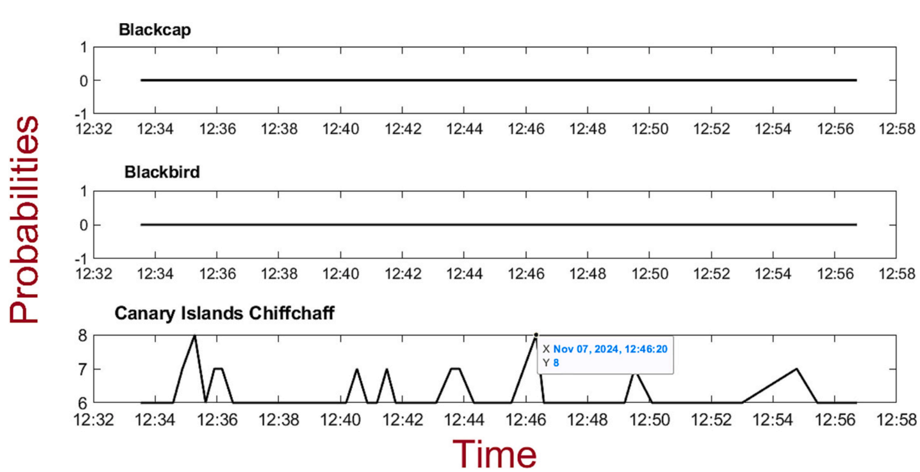

Fig. 11. Blackcap, Blackbird and Chiffchaff detections over a 50-min period at location 3 (rural area).

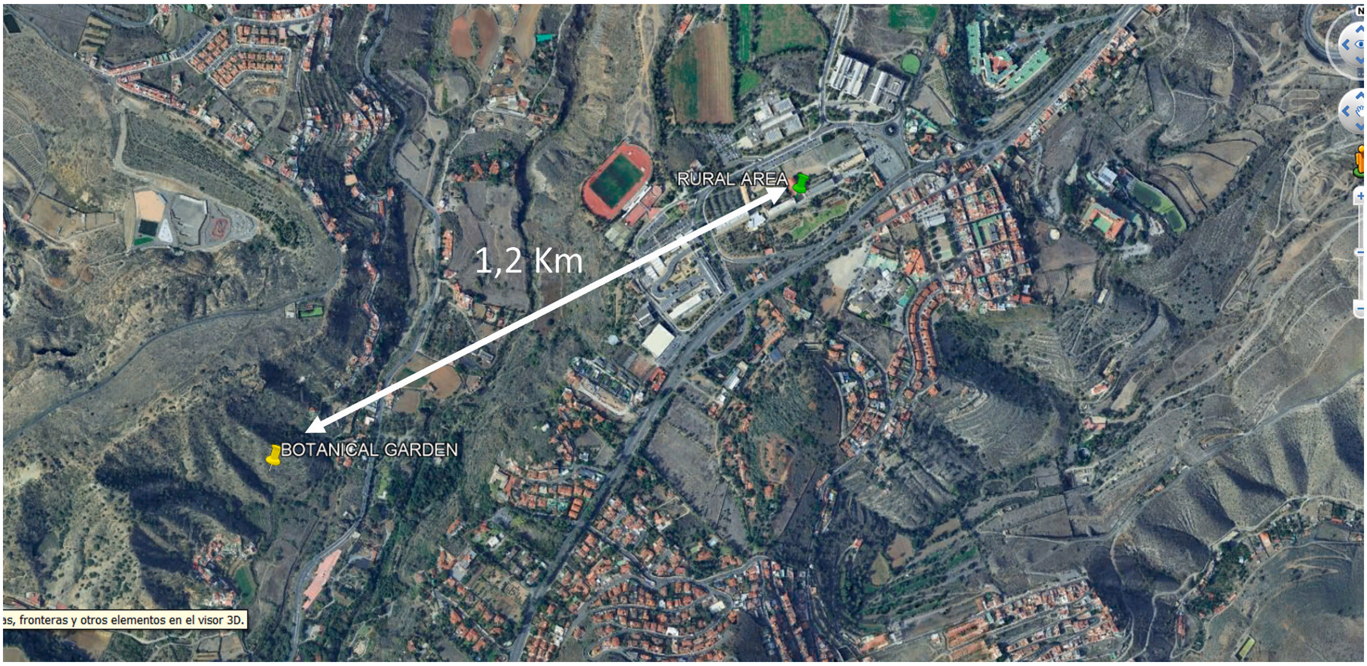

Fig. 12. Location of the BSD node (yellow) and the Gateway (green). LoRa link. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

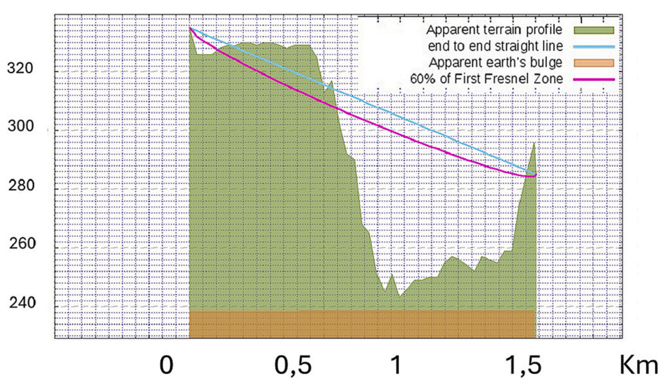

Fig. 13. The LoRa link profile of the tests in the locations shown in Fig. 12 obtained with the BOT-RF tool. dB versus Km.

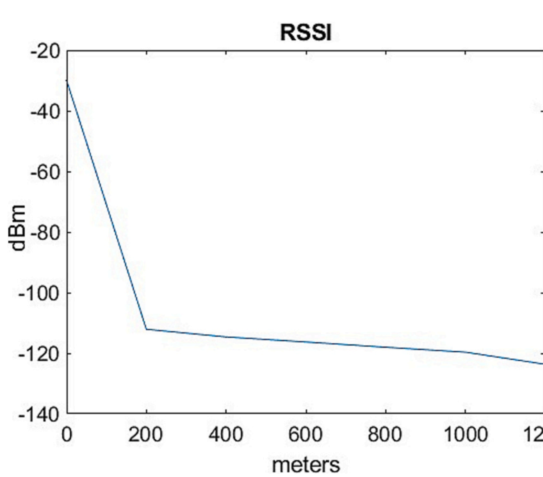

Fig. 14. RSSI of packets received at the gateway.

accurate ranking among the five classes.

Table 2 shows the metric by classes and global metrics of the second ranking system. It shows a very high performance, with overall accuracy of 98.8 % and a macro-average F1-score of 0.962. The Blackcap class has the relatively lowest performance, especially in terms of recall, while the other classes show a very accurate ranking.

3.1. Real scenarios detections

Using these networks in the corresponding measurement nodes, examples of the results of the measurements taken in these real scenarios over periods of time of approximately 1 h are shown. Fig. 9 details the presence detections in scenario 1 from 12:30 to 13:20 on 7 Nov 2024. What is shown is the output activation signal of the neural network, which, based on a threshold, triggers the detection alarm for the presence of the species in question. Fig. 10 shows the percentage of total detections over a 10-day testing period in the same scenario. Fig. 11 shows the results of the detection measurements in scenario 3 of Fig. 5 (rural area). It is important to highlight the robustness of the System in these scenarios, despite the moderate wind conditions in the case of the rural area and urban noise in the case of the other two.

3.2. LoRa communication system measurements

One of the test environments for the case of LoRa communication is the one shown in Fig. 12, which shows the location of the measurement point with respect to the Gateway. The distance between the two points

Table 1
Partial and total metrics obtained in Urban Park.

	Type of class	Accuracy:	Recall	F1-Score
Metrics by Class	Background	97.90 %	100 %	98.90 %
	Parakeet	99,50 %	100 %	99.80 %
	Blackcap	96.40 %	94.20 %	95.30 %
	Sparrow	96.60 %	94.10 %	95.30 %
	Blackbird	99.00 %	99.00 %	99.00 %
	Total Accuracy		Macro-Average F1-Score	
Global Metrics	99.00 %		97.70 %	

Table 2
Partial and total metrics obtained in Rural Area.

	Type of class	Accuracy:	Recall	F1-Score
Metrics by Class	Background	99.20 %	100 %	99.60 %
	Blackcap	97,70 %	81.7 %	89.00 %
	Blackbird	97.00 %	99.00 %	98.00 %
	Chiffchaff	99.00 %	97,70 %	98.30 %
Global Metrics	Total Accuracy 98.80 %		Macro-Average F1-Score 96.20 %	

is 1,2 Km. The gateway antenna was located at a height of 12 m while the portable equipment is shown in Fig. 6. The objective was to test the viability of this type of link despite the unevenness found, which is reflected in Fig. 13. To this extent, the BOT-RF application (Zennaro et al., 2017), which is a BOT of the "*Telegram*" application, was used to verify the feasibility of wireless links. It uses the database generated by the SRTM project that contains globe elevation data. The radiation model he uses to perform the simulations is the Longley-Rice Irregular Terrain Model (L-R ITM) (Longley and Rice, 1968).

Once simulated and calculated the Free Space Path Loss (FSPL =8 9.8 dB's) using the above tool, the RSSI measurements were made every 200 m between the two locations through fieldwork and the result is shown in Fig. 14. It was obtained using a Spreading Factor (SF) of 12 (Vangelista, 2017) and the maximum power of the Heltec LoRa Wireless Bridge SX1276 transmitter which is 20 dBm. With these levels, although there is a decrease in the bitrate, the connection is achievable, because the sensitivity of the receiver is −148 dBm. A transmission of the detections was made in the scenario and the level of RSSI received at the Gateway for each site is shown. Considering that packets received at a distance of 0,5 m arrive with a RSSI of −30 dBm, Eq. (5) is used to calculate the RSSI at each point of the journey. As can be seen, the measurements do not differ much from those calculated from the simulation.

$$RSSI_{rx} = RSSI_{0,5m} - FSPL_{Tx-Rx} \approx -30dBm - 89,8dB = -119,8dBm \tag{5}$$

4. Discussion

Despite using a low-cost mobile device with relatively limited processing capacity, along with accessories such as microphones or LoRa transceivers, the successful deployment of nodes has been demonstrated in two distinct environments. On the one hand, in urban areas (e.g., parks), where Wi-Fi access points are readily available, measurements of species presence probability have been conducted with highly acceptable accuracy, despite high noise pollution from conversations, traffic, and construction. On the other hand, the system's viability in more remote (rural) areas has also been validated through LPWAN transmission, which does not require 4G or 5G coverage and can reach several kilometers without repeaters, activating only in the absence of an access point.

It is important to note that not all possible species were considered in the system's evaluation, as the primary goal was to demonstrate its overall feasibility. However, with further training, both the success rate and the number of detectable species can be increased.

The evaluation results demonstrate promising performance across various metrics. High overall accuracy and macro-averaged F1-scores indicate the model's ability to effectively classify bird vocalizations while minimizing misclassifications. The high accuracy observed in the "Background" class suggests that the model effectively discriminates between environmental sounds and bird vocalizations, minimizing false alarms.

Class-specific metrics provide valuable insights into the model's performance for individual species. For example, while the model achieved high accuracy and recall for species such as "Parrot" and "Blackbird," lower recall values were observed for certain species like "Blackcap," indicating potential challenges in accurately detecting these species within the acoustic recordings. These findings highlight the importance of further investigation and potential improvements in the model's ability to accurately detect and classify less prevalent or acoustically challenging species.

The results obtained in the "Urban Park" and "Rural Area" environments provide valuable insights into the model's robustness across different acoustic landscapes. While the model demonstrated consistently high performance in both environments, further analysis is required to understand the factors influencing its performance in different acoustic contexts, such as noise levels, species diversity, and environmental complexity.

The main advantages of this architecture stem from its ability to operate with a lower bitrate compared to systems such as those proposed by other studies that require transmitting the audio fragment (Rovithis et al., 2021; Sullivan et al., 2009). This is achieved by directly transmitting the detection and classification results, which are processed in real-time at the EDGE layer of the architecture. The designed LoRa frame format includes information on the node's location, identity, and detected species in real-time, making the payload size used by LoRa more than sufficient.

Other studies perform real-time processing directly on the nodes (Velasco-Montero et al., 2024), but they focus on images, despite using a detection and classification approach similar to that of this work. However, their main limitation is that they require a Wi-Fi access point, as they do not consider the use of LPWAN technologies. Table 3 presents a comparison between various studies cited in the state of the art.

In summary, as shown, this work is the only one that enables real-time processing while eliminating the need for a high bit rate to transmit audio, as the processing is performed directly on the sensor node. Additionally, it does not require mobile network coverage, as it utilizes LoRa as the communication technology, achieving transmission ranges of more than ten kilometers. Another key advantage of this approach is its feasibility for deployment in remote areas without network coverage, as well as its ability to support a large number of sensor nodes due to its low cost.

At the same time, long-duration sound segments of each species have been used to cut them into only those of interest of 1 s. duration for the

Table 3
Comparison between different similar studies in terms of connectivity, latency and where they perform the processing.

Study	Architecture	Need for coverage	Communications Technology	Real time
LeBien et al. (2020)	Pipeline. Soundscape recordings previously collected	No	None	No
Sullivan et al. (2009)	Based in mobile recordings	Yes	4G	No
Rovithis et al. (2021)	Based in mobile recordings	Yes	4G	No
Cinkler et al. (2021)	Pre-processing on sensor nodes and post-processing in the cloud	Yes	Wi-Fi	Yes
Velasco-Montero et al. (2024)	On-site processing with camera traps and sending alarms via Wi-Fi	Yes	Wi-Fi	Yes
This Work	On-site processing using microphone and Rasperry Pi and sending alarms via Wi-Fi or LoRa	Not in the case of LoRa up to 10 Km	Wi-Fi LoRa	Yes

training. To do this, a filtering algorithm has been used, followed by a detection of the existence of bird songs by means of RMS of the signal instead of doing it manually (Rocha et al., 2015; Selin et al., 2007), simplifying the data acquisition process.

This work is based on the application of new IoT-based communication techniques to enable access to remote rural environments. It allows for the deployment of a low-cost IoT network capable of mapping, in real time, the presence of specific species across a wide area. In this way, bird populations can be easily monitored through a web or mobile application. The main challenge faced was conducting field tests and ensuring the interoperability of the two communication techniques in order to build an equivalent system for both. We are currently implementing a similar system for Smart Cities, using another communication technology that is more resistant to obstacles, although with a shorter range.

5. Conclusions

The feasibility of implementing a low-cost sensor node for bird detection through song recognition with local processing (EDGE computing) has been demonstrated. This enables its integration into a flexible Internet of Things (IoT) architecture that combines two different communication technologies. The system serves as a valuable tool for biodiversity analysis by detecting the presence of various bird species continuously and in real-time. It allows for analysis based on factors such as seasonality, time of day, and responses to external influences, such as the presence of invasive species. For example, during testing, the system detected the impact of the Rose-ringed parakeet (*Psittacula krameri*) on the Heineken Eurasian Blackcap (*Sylvia atricapilla heineken*).

The flexibility of this architecture simplifies the deployment of sensor nodes, as it does not necessarily rely on an access point or mobile network coverage. Moreover, with minor modifications, the system could incorporate data from additional sensors, such as wind speed, temperature, humidity, or solar radiation. Additionally, this type of node facilitates the monitoring of remote or isolated areas by strategically placing LoRa Gateways in locations with network access.

The system achieves accuracy levels comparable to other state-of-the-art solutions (98.8 % in urban areas and 99 % in rural areas), despite being implemented on portable devices with limited computational capacity. The next step involves improved training based on a large and diverse dataset, as well as enabling the generation of a custom dataset using the developed nodes themselves. This would involve systems capable of storing or transmitting detected sound segments of each species — a line of research that is currently emerging within the group. The main limitation of this system lies in the processing capabilities of the devices used, which currently results in suboptimal accuracy. However, the low cost of these devices enables a broader deployment of sensor nodes. In future work, our team plans to conduct long-term testing with a greater variety of bird species, continuously integrating sensor data with environmental information.

Funding

The project was supported by funding from the Institute for Technological Development and Innovation in Communications (IDeTIC) at ULPGC.

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

CRediT authorship contribution statement

Francisco A. Delgado-Rajó: Writing – original draft, Validation, Resources, Methodology, Investigation, Conceptualization. **Carlos M. Travieso-Gonzalez:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgements

This work has been funded by the Institute for Technological Development and Innovation in Communications (IDeTIC). ULPGC. We also thank the El Muelle Shopping Center (Muelle de Sta. Catalina, s / n, 35008 Las Palmas de Gran Canaria, Las Palmas, Spain) for providing us with locations for the placement of the nodes in the measurement stage.

Data availability

https://github.com/pcorajo/Bird-Songs-Detection-System
The remaining data will be made available upon request.

References

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. PeerJ 1, e103. DOI: https://peerj.com/.

Akçay, H.G., Kabasakal, B., Aksu, D., Demir, N., Öz, M., Erdoğan, A., 2020. Automated bird counting with deep learning for regional bird distribution mapping. Animals 10 (7), 1207. https://doi.org/10.3390/ani10071207.

Andreassen, T., Surlykke, A., Hallam, J., 2014. Semi-automatic long-term acoustic surveying: a case study with bats. Eco. Inform. 21, 13–24. https://doi.org/10.1016/j.ecoinf.2013.12.010.

Andrewartha, H.G., Birch, L.C., 1954. The Distribution and Abundance of Animals. Chicago University Press.

Ariyanto, M., Haryanto, I., Setiawan, J.D., Munadi, M., Radityo, M.S., 2019. Real-time image processing method using raspberry pi for a Car model. In: 2019 6th International Conference on Electric Vehicular Technology (ICEVT). IEEE, pp. 46–51. https://doi.org/10.1109/ICEVT48285.2019.8993866a.

Arowolo, Modupe, Aaron, William, Kugbiyi, Adeniyi, Eteng, Ubi, Iloh, Divine, Aguma, Chimaoge, Olagunju, Adeyemi, 2024. Integrating AI enhanced remote sensing technologies with IOT networks for precision environmental monitoring and predicative ecosystem management. World J. Adv. Res. Rev. 23, 2156–2166. https://doi.org/10.30574/wjarr.2024.23.2.2573.

Bell, G., 2003. The interpretation of biological surveys. Proc. Biol. Sci. 270, 2531–2542. https://doi.org/10.1126/science.293.5539.2413.

Carvalho, S., Gomes, E.F., 2023. Automatic classification of bird sounds: using MFCC and Mel spectrogram features with deep learning. Vietnam J. Comp. Sci. 10 (1), 39–54. https://doi.org/10.1142/S2196888822500300.

Chandu, B., Munikoti, A., Murthy, K.S., Murthy, G., Nagaraj, C., 2020. Automated bird species identification using audio signal processing and neural networks. In: 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, pp. 1–5. https://doi.org/10.1109/AISP48273.2020.9073584.

Chaudhari, B.S., Zennaro, M., Borkar, S., 2020. LPWAN technologies: emerging application characteristics, requirements, and design considerations. Future Intern. 12 (3), 46. https://doi.org/10.3390/fi12030046.

Cinkler, T., Kristóf, N., Simon, C., Vida, R., Rajab, H., 2021. Two-phase sensor decision: machine-learning for bird sound recognition and vineyard protection. IEEE Sensors J. https://doi.org/10.1109/JSEN.2021.3134817.

Delgado-Rajo, F., Alvarado Ramírez, I., 2024. Hybrid architectures to improve coverage in remote areas and incorporate long-range LPWAN multi-hop IoT strategies. IntechOpen. https://doi.org/10.5772/intechopen.113328.

Delgado-Rajo, F., Melian-Segura, A., Guerra, V., Perez-Jimenez, R., Sanchez-Rodriguez, D., 2020. Hybrid RF/VLC network architecture for the internet of things. Sensors 20 (2), 478. https://doi.org/10.3390/s20020478.

European Commission, 2011. Automatic acoustic monitoring and inventorying of biodiversity (LIFE08 NAT/GR/000539). LIFE Programme. Retrieved from. https://webgate.ec.europa.eu/life/publicWebsite/project/LIFE08-NAT-GR-000539/automatic-acoustic-monitoring-and-inventorying-of-biodiversity.

Ferdoush, S., Li, X., 2014. Wireless sensor network system design using raspberry pi and Arduino for environmental monitoring applications. Proc. Comp. Sci. 34, 103–110. https://doi.org/10.1016/j.procs.2014.07.059.

García-Martín, J.P., Torralba, A., 2021. Model of a device-level combined wireless network based on NB-IoT and IEEE 802.15.4 standards for low-power applications. Sensors 21 (11), 3718. https://doi.org/10.3390/s21113718.

Goitia-Urdiain, Teresa Sauras-Yera, Llorente, Gustavo A., Pujol-Buxó, Eudald, 2024. Software-dependent biases in the recognition of di- and tri-syllabic bird songs can create false interpretations of bird abundance and singing activity. Ecol. Inform. 79, 102397. ISSN 1574-9541. https://doi.org/10.1016/j.ecoinf.2023.102397.

Guma, Ali, Mijwil, Maad, Adamopoulos, Ioannis, Namuq, Jenan, 2025. Leveraging the internet of things, remote sensing, and artificial intelligence for sustainable forest management. Babylonian J. Intern. Things 2025, 1–65. https://doi.org/10.58496/BJIoT/2025/001.

Guralnick, R., Walls, R., Jetz, W., 2017. Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling, and assessment. Ecography 41, 713–725. https://doi.org/10.1111/ecog.02942.

Islam, B., Valles, D., Hibbitts, T.J., Ryberg, W.A., Walkup, D.K., Forstner, M.R.J., 2023. Animal species recognition with deep convolutional neural networks from ecological camera trap images. Animals 13 (9), 1526. https://doi.org/10.3390/ani13091526.

Jančovič, P., Kküer, M., 2011. Automatic detection and recognition of tonal bird sounds in noisy environments. EURASIP J. Adv. Sign. Proc. 2011, 982936. https://doi.org/10.1155/2011/982936.

Jetz, W., McPherson, J.M., Guralnick, R.P., 2012. Integrating biodiversity distribution knowledge: toward a global map of life. Trends Ecol. Evol. 27 (2), 151–159. https://doi.org/10.1016/j.tree.2011.09.007.

Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Turak, E., 2019. Essential biodiversity variables for mapping and monitoring species populations. Nat. Ecol. Evol. 3 (4), 539–551. https://doi.org/10.1038/s41559-019-0826-1.

Kahl, Stefan, Wood, Connor M., Eibl, Maximilian, Klinck, Holger, 2021. BirdNET: A deep learning solution for avian diversity monitoring. Ecol. Inform. 61, 101236. ISSN 1574-9541. https://doi.org/10.1016/j.ecoinf.2021.

La Sorte, F.A., Lepczyk, C.A., Burnett, J.L., Hurlbert, A.H., Tingley, M.W., Zuckerberg, B., 2018. Opportunities and challenges for big data ornithology. Condor 120 (2), 414–426. https://doi.org/10.1650/CONDOR-17-206.1.

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J.P., Dodhia, R., Ferres, J.L., Aide, T. M., 2020. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Eco. Inform. 59, 101113. https://doi.org/10.1016/j.ecoinf.2020.101113.

Lees, A.C., Haskell, L., Allinson, T., Bezeng, S.B., Burfield, I.J., Renjifo, L.M., Butchart, S. H., 2022. State of the world's birds. Annu. Rev. Environ. Resour. 47, 231–260. https://doi.org/10.1146/annurev-environ-112420-.

Liechti, F., Aschwanden, J., Blew, J., Boos, M., Brabant, R., Dokter, A.M., Sapir, N., 2019. Cross-calibration of different radar systems for monitoring nocturnal bird migration across Europe and the near east. Ecography 42 (6), 887–898. https://doi.org/10.1111/ecog.04041.

Liu, Daidai, Xiao, Hanguang, Chen, Kai, 2025. Research progress in bird sounds recognition based on acoustic monitoring technology: A systematic review. Appl. Acoust. 228, 110285. ISSN 0003-682X. https://doi.org/10.1016/j.apacoust.2024.110285.

Longley, A.G., Rice, P.L., 1968. Prediction of Tropospheric Radio Transmission Loss over Irregular Terrain (ESSA Tech. Report ERL 79-ITS 67).. U.S. Government Printing Office.

McBride, W.J., Courter, J.R., 2019. Using Raspberry Pi microcomputers to remotely monitor birds and collect environmental data. Eco. Inform. 54, 101016. https://doi.org/10.1016/j.ecoinf.2019.101016.

Nayyar, A., Puri, V., 2016. A review of Arduino boards, Lilypad's & Arduino shields. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, pp. 1485–1492.

Nilsson, C., Dokter, A.M., Schmid, B., Scacco, M., Verlinden, L., Bäckman, J., Liechti, F., 2018. Field validation of radar systems for monitoring bird migration. J. Appl. Ecol. 55 (6), 2552–2564. https://doi.org/10.1111/1365-2664.13174.

Noumida, Rajan, R., 2021. Deep learning-based automatic bird species identification from isolated recordings. In: 2021 8th International Conference on Smart Computing and Communications (ICSCC), pp. 252–256. https://doi.org/10.1109/ICSCC51209.2021.9528234.

Pereira, E., et al., 2023. RFID technology for animal tracking: a survey. IEEE J. Radio Frequency Identific. 7, 609–620. https://doi.org/10.1109/JRFID.2023.3334952.

Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. J. Avian Biol. 49, jav-01447. https://doi.org/10.1111/jav.01447.

Rocha, L.H., Ferreira, L.S., Paula, B.C., Rodrigues, F.H., Sousa-Lima, R.S., 2015. An evaluation of manual and automated methods for detecting sounds of maned wolves (*Chrysocyon brachyurus* Illiger 1815). Bioacoustics 24, 185–198. https://doi.org/10.1080/09524622.2014.991448.

Rovithis, E., Moustakas, N., Vogklis, K., Drossos, K., Floros, A., 2021. Towards citizen science for smart cities: a framework for a collaborative game of bird call recognition based on internet of sound practices. ArXiv. https://doi.org/10.48550/arXiv.2103.16988 abs/2103.16988.

Schmeller, D.S., Henle, K., Loyau, A., Besnard, A., Henry, P.-Y., 2012. Bird-monitoring in Europe - a first overview of practices, motivations and aims. Nat. Conserv. 2, 41–57. https://doi.org/10.3897/natureconservation.2.3644.

Selin, A., Turunen, J., Tanttu, J.T., 2007. Wavelets in recognition of bird sounds. EURASIP J. Appl. Sign. Proc. 2007, 141. https://doi.org/10.1155/2007/141.

Stowell, D., Wood, M., Stylianou, Y., Glotin, H., 2016. Bird detection in audio: a survey and a challenge. In: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6. https://doi.org/10.1109/MLSP.2016.7738875.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142 (10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006.

The MathWorks Inc, 2024. MATLAB (Version R2024a). The MathWorks, Inc, Natick, Massachusetts.

Travieso, C., Noda, J., Sánchez-Rodríguez, D., 2021. Acoustic identification of insects based on cepstral data fusion and hidden Markov models. In B978-0-12-815160-0.00018-9. https://doi.org/10.1016/B978-0-12-815160-0.00018-9.

Vangelista, L., 2017. Frequency shift chirp modulation: the LoRa modulation. IEEE Sign. Proc. Lett. 24 (12), 1818–1821. https://doi.org/10.1109/LSP.2017.2762960.

Velasco-Montero, Delia, Fernández-Berni, Jorge, Carmona-Galán, Ricardo, Sanglas, Ariadna, Palomares, Francisco, 2024. Reliable and efficient integration of AI into camera traps for smart wildlife monitoring based on continual learning. Ecol. Inform. 83, 102815. ISSN 1574-9541. https://doi.org/10.1016/j.ecoinf.2024.102815.

Ventura, T.M., Oliveira, A.G., Ganchev, T.D., Figueiredo, J.M., Jahn, O., Marques, M.I., Schuchmann, K.-L., 2015. Audio parameterization with robust frame selection for improved bird identification. Expert Syst. Appl. 42 (22), 8463–8471. https://doi.org/10.1016/j.eswa.2015.07.002.

Vidaña-Vila, Ester, Navarro, Joan, Alsina-Pagès, Rosa Ma, Ramírez, Álvaro, 2020. A two-stage approach to automatically detect and classify woodpecker (Fam. Picidae) sounds. Appl. Acoust. 166, 107312. ISSN 0003-682X. https://doi.org/10.1016/j.apacoust.2020.107312.

Weychan, R., Marciniak, T., Dabrowski, A., 2015, September. Implementation aspects of speaker recognition using Python language and raspberry Pi platform. In: 2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA). IEEE, pp. 162–167. https://doi.org/10.1109/SPA.2015.7365153.

Wielgat, R., Świętojański, P., Potempa, T., Lisowska-Lis, A., Król, D., 2011. Detection of bird species using prefiltering and hidden Markov models. EURASIP J. Adv. Sign. Proc. 2011 (1), 982936.

xeno-canto, 2020. Sharing Bird Sounds from Around the World. Retrieved May 13, 2020, from. https://www.xeno-canto.org/.

Youngblood, M., 2019. A raspberry pi-based, RFID-equipped birdfeeder for the remote monitoring of wild bird populations. Ringing Migr. 34 (1), 25–32. https://doi.org/10.1080/03078698.2019.1759908.

Zennaro, M., Rainone, M., Pietrosemoli, E., 2017. Radio link planning made easy with a telegram bot. In: Gaggi, O., Manzoni, P., Palazzi, C., Bujari, A., Marquez-Barja, J. (Eds.), Smart Objects and Technologies for Social Good. GOODTECHS 2016, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 195. Springer, Cham. https://doi.org/10.1007/978-3-319-61949-1_31.